

僵尸网络域名家族的检测方法、装置、设备及存储介质

本申请要求于 2018 年 12 月 24 日提交中国专利局、申请号为 201811584694.2、发明名称为“僵尸网络域名家族的检测方法、装置、设备及存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本
5 申请中。

技术领域

本申请涉及信息安全技术领域，特别是涉及一种僵尸网络域名家族的检测方法、装置、设备以及计算机可读存储介质。

10 背景技术

僵尸网络对网络安全构成严重威胁，不法分子通过僵尸网络发起分布式拒绝服务攻击（Distributed Denial of Service, DDoS）、进行恶意挖矿、信息窃取、发送垃圾邮件等，严重危及国家、企业、组织和个人的利益，快速、准确的识别僵尸网络的通信并及时阻断有着非常重大的意义。大量的
15 的僵尸网络基于 DNS 协议发送命令与控制（Command and Control, C&C）信息与僵尸主机通信。

主流的僵尸网络域名家族检测方案主要包括基于文法特征的检测与基于病毒流量的检测两种。

基于文法特征的僵尸网络域名家族检测认为属于同一僵尸网络家族的
20 域名具备文法特征的相似性。同一僵尸网络家族的域名往往使用相同的域名生成算法（Domain Generation Algorithm, DGA）生成，通过生成大量的具有相似文法特征的随机域名来逃脱黑名单的检测。通过提取例如域名中辅音字母占的比例、最长有意义的长度、n-gram（文字或语言中的 n 个连续的单词组成序列的分布）等特征，可以将由不同 DGA 算法生成的域名
25 区分开，从而发现属于同一僵尸网络家族的域名。最近，循环神经网络（Recurrent Neural Network, RNN）也被用于 DGA 域名的检测中，通过 RNN 学习出域名构成的字符序列的特征，发现属于同一僵尸网络家族的 DGA 域名。然而，并不是所有的僵尸网络家族都只使用 DGA 算法来生成

C&C 域名，一旦同一家族的 C&C 域名之间不具备明显的文法特征上的相似性，这一类的僵尸网络域名家族检测算法就不能获得很好的性能。

基于病毒流量的僵尸网络域名家族检测通过恶意文件的家族信息对 C&C 域名按家族进行聚类并检测。通过分析属于同一病毒家族的恶意文件访问的域名情况可以分析出两个域名是否属于同一个僵尸网络家族。虽然这种方法并不只是简单的依赖于域名的文法特征，但是会受到病毒样本的数目的制约。此外，病毒会通过访问一些合法的域名对这种方法造成干扰，导致一定误报的产生。一些研究表明，大量 C&C 域名的活跃时间会比对应的病毒样本被获得的时间早上数周，甚至数月，这就导致基于病毒流量的方法具有一定的滞后性，不能第一时间发现威胁，排除威胁。

发明内容

本申请的目的是提供一种僵尸网络域名家族的检测方法、装置、设备以及计算机可读存储介质，以解决现有僵尸网络域名家族检测中检测维度单一、过于依赖病毒样本收集、检测实时性较差的问题。

为解决上述技术问题，本申请提供一种僵尸网络域名家族的检测方法，包括：

获取可疑域名；

基于各所述可疑域名在不同维度之间的关联性，构建域名时空关联图；在所述域名时空关联图中，各可疑域名作为一个节点，拥有至少一种关联性的两个域名之间构成一条边，两个域名之间的关联性作为边的属性值；

根据图计算中各个节点紧凑程度的判断指标，确定得到所述域名时空关联图中联系紧凑的域名，将对应域名的集合作为僵尸网络域名家族。

可选地，所述根据图计算中各个节点紧凑程度的判断指标，确定得到所述域名时空关联图中联系紧凑的域名包括：

根据稀疏度和平均聚类程度，确定得到所述域名时空关联图中联系紧凑的域名。

可选地，所述根据图计算中各个节点紧凑程度的判断指标，确定得到所述域名时空关联图中联系紧凑的域名包括：

对所述域名时空关联图中的异常子图以及异常节点进行去除，确定得到所述域名时空关联图中联系紧凑的域名。

可选地，所述对所述域名时空关联图中的异常子图以及异常节点进行去除包括：

- 5 依据连通性将所述域名时空关联图分解为多个连通分支构成的子图，去除不满足预设第一连通性指标的异常子图；

使用社区发现算法对去除异常子图后剩余的子图进行分割，得到多个连通分支构成的社区，去除不满足预设第二连通指标的异常社区；

- 10 采用网页排名算法度量每个节点在去除异常社区后剩余的社区中的重要程度，去除重要程度低的潜在异常点。

可选地，所述不同维度之间的关联性包括以下特征中的任意组合：

域名文法特征上的相似性、域名与病毒关联上的相似性、域名解析出的IP地址上的相似性、域名访问流量上的相似性。

可选地，在所述将对应域名的集合作为僵尸网络域名家族之后还包括：

- 15 对所述僵尸网络域名家族的家族特征进行分析，并根据分析得到的家族特征对应用场景进行安全检测。

可选地，在所述将对应域名的集合作为僵尸网络域名家族之后还包括：

通过判断新域名与已知域名之间的关联性，确定新域名是否属于所述僵尸网络域名家族，以便对所述僵尸网络域名家族的变种和扩张进行监控。

- 20 本申请还提供了一种僵尸网络域名家族的检测装置，包括：

域名获取模块，用于获取可疑域名；

关联图构建模块，用于基于各所述可疑域名在不同维度之间的关联性，构建域名时空关联图；在所述域名时空关联图中，各可疑域名作为一个节点，拥有至少一种关联性的两个域名之间构成一条边，两个域名之间的关
25 联性作为边的属性值；

检测模块，用于根据图计算中各个节点紧凑程度的判断指标，确定得到所述域名时空关联图中联系紧凑的域名，将对应域名的集合作为僵尸网络域名家族。

可选地，所述检测模块用于：根据稀疏度和平均聚类程度，确定得到

所述域名时空关联图中联系紧凑的域名。

可选地，所述检测模块用于：对所述域名时空关联图中的异常子图以及异常节点进行去除，确定得到所述域名时空关联图中联系紧凑的域名。

- 5 可选地，所述检测模块用于：依据连通性将所述域名时空关联图分解为多个连通分支构成的子图，去除不满足预设第一连通性指标的异常子图；使用社区发现算法对去除异常子图后剩余的子图进行分割，得到多个连通分支构成的社区，去除不满足预设第二连通指标的异常社区；采用网页排名算法度量每个节点在去除异常社区后剩余的社区中的重要程度，去除重要程度低的潜在异常点。

- 10 可选地，所述不同维度之间的关联性包括以下特征中的任意组合：域名文法特征上的相似性、域名与病毒关联上的相似性、域名解析出的 IP 地址上的相似性、域名访问流量上的相似性。

可选地，还包括：

- 15 安全检测模块，用于在将对应域名的集合作为僵尸网络域名家族之后，对所述僵尸网络域名家族的家族特征进行分析，并根据分析得到的家族特征对应用场景进行安全检测。

可选地，还包括：

- 20 监控模块，用于在将对应域名的集合作为僵尸网络域名家族之后，通过判断新域名与已知域名之间的关联性，确定新域名是否属于所述僵尸网络域名家族，以便对所述僵尸网络域名家族的变种和扩张进行监控。

本申请还提供了一种僵尸网络域名家族的检测设备，包括：

存储器，用于存储计算机程序；

处理器，用于执行所述计算机程序时实现上述任一种僵尸网络域名家族的检测方法的步骤。

- 25 本申请还提供了一种计算机可读存储介质，所述计算机可读存储介质上存储有计算机程序，所述计算机程序被处理器执行时实现上述任一种所述僵尸网络域名家族的检测方法的步骤。

本申请所提供的僵尸网络域名家族的检测方法，通过获取可疑域名；基于各可疑域名在不同维度之间的关联性，构建域名时空关联图；在域名

时空关联图中，各可疑域名作为一个节点，拥有至少一种关联性的两个域名之间构成一条边，两个域名之间的关联性作为边的属性值；根据图计算中各个节点紧凑程度的判断指标，确定得到域名时空关联图中联系紧凑的域名，将对应域名的集合作为僵尸网络域名家族。本申请与基于文法特征的家族检测方法只关心某一层面的关联性相比，能够将域名之间各个不同维度之间的关联性统一用关联图的形式表示出来，因而具备更强的检测能力。同时，本申请的检测不依赖于病毒流量，能够更快速检测出僵尸域名家族，并及时进行处理，降低造成的各种损失，具备更广泛的适用性。此外，本申请还提供了具有上述技术优点的僵尸网络域名家族的检测装置、设备以及计算机可读存储介质。

附图说明

为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为本申请所提供的僵尸网络域名家族的检测方法的一种具体实施方式的流程图；

图 2 为本申请对所述域名时空关联图中的异常子图以及异常节点进行去除的过程流程图；

图 3 为本申请实施例提供的僵尸网络域名家族的检测装置的结构框图。

具体实施方式

为了使本技术领域的人员更好地理解本申请方案，下面结合附图和具体实施方式对本申请作进一步的详细说明。显然，所描述的实施例仅仅是

本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

本申请所提供的僵尸网络域名家族的检测方法的一种具体实施方式的流程图如图 1 所示，该方法包括：

步骤 S101：获取可疑域名。

可疑域名指的是排除了明显为正常的合法域名，并且检测出具备至少一项异常行为的域名，如域名文法特征特别像 DGA 域名、活跃时间总是集中在凌晨等。排除明显为正常的域名可以通过白名单技术，如将 Alexa 排名较高的域名加入白名单，并且认为这些排名较高的域名不太可能成为僵尸网络的 C&C 域名。

步骤 S102：基于各所述可疑域名在不同维度之间的关联性，构建域名时空关联图；在所述域名时空关联图中，各可疑域名作为一个节点，拥有至少一种关联性的两个域名之间构成一条边，两个域名之间的关联性作为边的属性值。

在步骤 S101 获取到可疑域名之后，可以对这些可疑域名进行分析，能够得到各个可疑域名在不同维度之间的关联性。作为一种具体实施方式，可以将域名之间的关联性表示为三元组的形式，如（域名 1，域名 2，关联性指标）。其中，三元组中的关联性指标包含关联的维度与在该维度上两个域名的相似性度量。针对不同的关联维度，可以设置针对性的不同的度量指标，用以衡量两个域名在该维度上的相似程度。两个域名可以在多个特征维度上具有相似性，表示在三元组中，关联性指标部分可以用数组来表示。

其中，所述不同维度之间的关联性包括以下特征中的任意组合：域名文法特征上的相似性、域名与病毒关联上的相似性、域名解析出的 IP 地址上的相似性、域名访问流量上的相似性。当然，并不限于上述这几种，可以选取尽可能多、尽可能全面的维度对域名的关联性进行分析，包罗域名行为的各个方面，并且如果发现新的维度上的关联性，就可以将该维度的特征分析加入到分析中来，具备很强的可扩展性。

本实施例中，可以使用图数据库基于三元组进行时空关联图的构建与存储，后续随着时间的推进对时空关联图的扩充也可以很方便的基于图数据库进行操作。域名时空关联图刻画了域名之间在时间与空间上的关联性，空间上的关联即为各个维度上的关联性，时间上的关联指的是域名之间随着时间的变化关系，两个原本并没有关联的域名随着时间的推进以及某些特定安全事件的产生，有可能会在某些维度产生关联。

在构建的域名时空关联图中，各可疑域名作为一个节点，拥有至少一种关联性的两个域名之间构成一条边，两个域名之间的关联性作为边的属性值。一条边可以有多个属性，每个属性对应着一个维度的关联性，可以包括域名被检测可疑的时间、域名能否成功解析出 IP 地址等。例如，域名 1 与域名 2 之间具有关联性，则在域名时空关联图中，用一条边将域名 1 与域名 2 连起来。也就是说，只有在至少一个维度上存在关联的域名才会存在相连的边。将域名之间的关联性，例如（关联维度 A，关联维度 A 对应的相似程度；关联维度 B，关联维度 B 对应的相似程度；关联维度 C，关联维度 C 对应的相似程度……）作为边的属性。

步骤 S103：根据图计算中各个节点紧凑程度的判断指标，确定得到所述域名时空关联图中联系紧凑的域名，将对应域名的集合作为僵尸网络域名家族。

连通性，也就是各个节点紧凑程度，可以有多个度量标准，不同的度量标准在不同的场景下拥有更好的性能。稀疏度和平均聚类程度是两个比较合适的度量子图连通性的指标。稀疏度定义为图中边的数目与图中节点构成的全连接图边的数目的比值；平均聚类程度定义为节点周边三角形的数目与潜在三角形数目比值的平均。稀疏度指标在节点数目较少时能够准确衡量图的紧凑程度，而平均聚类程度则在节点数目较多的时候能够准确衡量图的紧凑程度。因此，稀疏度与平均聚类程度配合使用能够更加准确的衡量图的紧凑程度或连通程度。作为一种具体实施方式，本申请可以根据稀疏度和平均聚类程度，确定得到所述域名时空关联图中联系紧凑的域名。

本申请所提供的僵尸网络域名家族的检测方法，通过获取可疑域名；

基于各可疑域名在不同维度之间的关联性，构建域名时空关联图；在域名时空关联图中，各可疑域名作为一个节点，拥有至少一种关联性的两个域名之间构成一条边，两个域名之间的关联性作为边的属性值；根据图计算中各个节点紧凑程度的判断指标，确定得到域名时空关联图中联系紧凑的域名，将对应域名的集合作为僵尸网络域名家族。本申请与基于文法特征的家族检测方法只关心某一层面的关联性相比，能够将域名之间各个不同维度之间的关联性统一用关联图的形式表示出来，因而具备更强的检测能力。同时，本申请的检测不依赖于病毒流量，能够更快速检测出僵尸域名家族，并及时进行处理，降低造成的各种损失，具备更广泛的适用性。

10

本申请能够综合考虑域名之间各个维度的相似度，构建出域名之间的时空关联图，基于图计算排除图中的异常子图，并排除正常子图中的异常节点，只保留联系最紧密的域名构成的僵尸网络家族。最后，以僵尸网络家族或紧密子图为单位，从病毒流量、域名文法特征等维度对僵尸网络家族进行进一步的分析，排除可能的误报。

本申请实施例中，根据图计算中各个节点紧凑程度的判断指标，确定得到所述域名时空关联图中联系紧凑的域名的过程可以具体为：对所述域名时空关联图中的异常子图以及异常节点进行去除，确定得到所述域名时空关联图中联系紧凑的域名。参照图 2，本申请对所述域名时空关联图中的异常子图以及异常节点进行去除的过程可以具体包括：

步骤 S201：依据连通性将所述域名时空关联图分解为多个连通分支构成的子图，去除不满足预设第一连通性指标的异常子图；

其中，预设第一连通性指标可以为子图中的节点数目大于预设阈值并且连通性强。当子图中的节点数目过少时，说明对于该家族的僵尸网络域名收集的信息过少，因此很难准确分析出该家族域名之间的关联性，于是将这些节点数目较少的子图丢弃，等到收集到足够的信息后再做进一步的分析。这样，可以将子图中节点数目足够多并且连通性较强的子图进行进一步分析，其余的子图则丢弃。

需要指出的是，本步骤中的连通性不只包含某一个特定的度量指标，

也可以包含多个图的紧凑程度度量指标的配合使用。

- 本步骤中，由于初始的时空关联图包含需要进行分析的所有僵尸网络家族，在进行连通分支发现时需要处理海量的以图结构存储的性质，对内存以及计算能力的需求比较大，因此需要分布式的处理。在计算连通分支构成的子图的连通度时，可以根据需求选取一个或多个度量指标，将连通不够紧密的子图排除。

步骤 S202：使用社区发现算法对去除异常子图后剩余的子图进行分割，得到多个连通分支构成的社区，去除不满足预设第二连通指标的异常社区；

- 10 社区指的是一种内部连接紧密，而外部稀疏的群体结构。社区发现算法能够通过衡量节点之间的关联发现这样的群体结构。使用社区发现算法将一个连通性较强的分支构成的子图分割成若干个连通性更强的社区，丢弃那些连通性不够强的社区。通过社区发现算法获得的子图具有更强的紧凑程度，因此属于同一个僵尸网络家族的可能就更大，出现误报的可能就
- 15 更小。

需要指出的是，本申请中第二连通指标比第一连通性指标在连通性指标上应该更加严格，因为社区构成的子图的紧凑程度与该社区所在的连通分支构成的子图的连通程度相比更强。

- 本步骤中在选择社区发现算法时，需要综合考虑算法的精度与时间复杂度。一些具有很好社区划分性能的计算资源，耗费大量的计算时间；而一些启发式的算法虽然能够快速地进行社区发现，但划分的结果有待进一步考虑，因此在实际部署时，需要根据实际的需求对二者进行权衡。
- 20

- 步骤 S203：采用网页排名算法度量每个节点在去除异常社区后剩余的社区中的重要程度，去除重要程度低的潜在异常点。
- 25

网页排名算法最早用来衡量特定网页相对于搜索引擎索引中的其他网页而言的重要程度，在一个图中网页排名算法能够根据一个节点与其他节点的关联程度衡量出该节点在整个图中的重要性。对于连通性足够强的社区，使用网页排名算法度量每个节点在社区中的重要程度，排除重要程度

较低的潜在异常点，进一步提升社区的紧凑程度与连通性。

网页排名算法输入一张图，输出图中每个节点在图中的重要程度的度量，所有节点的网页排名值之和为 1。直观的，如果一个节点与越多的其他节点有关联，其在图中的重要性就越强。把图中的节点按照网页排名值进行排序，可以发现那些在图中重要性较低的与图中其他节点关联性较低的节点。通过将这些节点从图中删去，可以保证图中的节点彼此之间均拥有较强的连通性。基于网页排名值进行异常点检测的方法可以包含多种统计学方法，通过对图中节点 page rank 值的分布进行分析，可以发现那些网页排名值明确低于平均水平的节点，然后就可以将这些节点从图中删去。

10 属于同一个社区的域名已经经过了之前多次的过滤，保留的节点之间拥有很强的关联性，具有很强的内聚性，因此认为这些节点属于同一个僵尸网络家族，至此就完成同一僵尸网络家族节点的聚类。

在上述任一实施例的基础上，本申请所提供的僵尸网络域名家族的检测方法在所述将对应域名的集合作为僵尸网络域名家族之后还包括：

对所述僵尸网络域名家族的家族特征进行分析，并根据分析得到的家族特征对应用场景进行安全检测。

20 各个家族的主要特征主要由该家族域名构成的社区中边的属性值决定，边的属性值包括域名文法特征上的相似性、域名与病毒关联上的相似性、域名解析出的 IP 地址上的相似性、域名访问流量上的相似性等方面中的一个或多个。通过对边的属性值进行分析能够了解到该家族的域名是通过什么原因而聚类到一起，从而了解到该僵尸网络家族中域名主要的特征是什么。

25 产出的黑域名可被部署在多种安全检测的应用场景，包括但不限于加入黑域名库、进行域名信誉评分等。被聚类在同一个僵尸网络家族的域名中通常既包含已知的 C&C 域名，也包含新发现的域名，通过评测新发现域名与已知 C&C 域名的关联程度，可以给出这些新发现域名为恶意的置信度，即确定域名信誉。

进一步地，本申请还可以包括：通过判断新域名与已知域名之间的关

联性，确定新域名是否属于所述僵尸网络域名家族，以便对所述僵尸网络域名家族的变种和扩张进行监控。

随着域名信息收集的越来越完善，将会获得更多关于某个僵尸网络家族的域名信息与域名关联信息，该家族的域名时空关联图也会变得越来越完善。原本很多因为图中节点数目不足够或者连通程度不够紧密而未被继续分析的域名家族随着信息的收集，能够不再被过滤掉，从而产出更多的检测结果；原本被认为是属于两个僵尸网络家族的子图也有可能因为信息收集的不断完善而被发现是属于同一个僵尸网络家族。

黑客为了逃脱检测会使用新的域名进行通信，由于能够在多个维度评测一个新域名与已知域名的关联性，并统一用图的数据结构表示，很容易从僵尸网络家族的时空关联图中判断新域名是否属于之前已经检测出的僵尸网络家族，从而观察到该家族的演变与扩张的过程。此外，由于基于图计算的僵尸网络域名家族检测方法的实时性，能够在第一时间检测出某个僵尸网络家族新增的 C&C 域名。

本申请通过对域名之间各个维度的相似性进行分析，并将不同维度的相似性统一用图的数据结构进行表示，构成域名的时空关联图。在时空关联图上执行图计算算法，经过连通分支发现、社区发现、基于网页排名的异常点检测发现具有强内聚、弱外联性质的若干子图，每个子图对应一个僵尸网络域名家族。本申请并不只是仅仅依赖某一维度的特征对属于同一僵尸网络家族的 C&C 域名进行聚类与检测，而是综合考虑域名之间各个维度的相似性关系，因此能够对属于同一僵尸网络家族的域名更加全面的分析，并且具备更强的可扩展性，新增的维度的特征也可以很容易加入到本申请提出的检测方法中去。本申请提出的检测方法具备很强的实时性，并不需要一定要获得 C&C 域名对应的恶意文件样本之后才能对其所属的僵尸网络家族进行分析。

进一步地，本申请能够基于家族可以分析不同类型 C&C 域名不同的性质并将家族分析结果进一步应用到其他的安全检测场景中。并且，还能够跟踪僵尸网络域名家族的演变与扩张，第一时间发现家族中新增的 C&C 域名。

下面对本申请实施例提供的僵尸网络域名家族的检测装置进行介绍，下文描述的僵尸网络域名家族的检测装置与上文描述的僵尸网络域名家族的检测方法可相互对应参照。

- 5 图 3 为本申请实施例提供的僵尸网络域名家族的检测装置的结构框图，参照图 3 僵尸网络域名家族的检测装置可以包括：

域名获取模块 100，用于获取可疑域名；

- 10 关联图构建模块 200，用于基于各所述可疑域名在不同维度之间的关联性，构建域名时空关联图；在所述域名时空关联图中，各可疑域名作为一个节点，拥有至少一种关联性的两个域名之间构成一条边，两个域名之间的关联性作为边的属性值；

检测模块 300，用于根据图计算中各个节点紧凑程度的判断指标，确定得到所述域名时空关联图中联系紧凑的域名，将对应域名的集合作为僵尸网络域名家族。

- 15 其中，所述检测模块 300 具体用于：根据稀疏度和平均聚类程度，确定得到所述域名时空关联图中联系紧凑的域名。

作为一种具体实施方式，所述检测模块 300 具体用于：对所述域名时空关联图中的异常子图以及异常节点进行去除，确定得到所述域名时空关联图中联系紧凑的域名。

- 20 其中，所述不同维度之间的关联性包括以下特征中的任意组合：

域名文法特征上的相似性、域名与病毒关联上的相似性、域名解析出的 IP 地址上的相似性、域名访问流量上的相似性。

在上述任一实施例的基础上，本申请所提供的僵尸网络域名家族的检测装置还可以进一步包括：

- 25 安全检测模块，用于在将对应域名的集合作为僵尸网络域名家族之后，对所述僵尸网络域名家族的家族特征进行分析，并根据分析得到的家族特征对应用场景进行安全检测。

此外，在上述任一实施例的基础上，本申请所提供的僵尸网络域名家族的检测装置还可以进一步包括：

监控模块，用于在将对应域名的集合作为僵尸网络域名家族之后，通过判断新域名与已知域名之间的关联性，确定新域名是否属于所述僵尸网络域名家族，以便对所述僵尸网络域名家族的变种和扩张进行监控。

本实施例的僵尸网络域名家族的检测装置用于实现前述的僵尸网络域名家族的检测方法，因此僵尸网络域名家族的检测装置中的具体实施方式可见前文中的僵尸网络域名家族的检测方法的实施例部分，例如，域名获取模块 100，关联图构建模块 200，检测模块 300，分别用于实现上述僵尸网络域名家族的检测方法中步骤 S101，S102，S103，所以，其具体实施方式可以参照相应的各个部分实施例的描述，在此不再赘述。

10 另一方面，本申请还提供了一种僵尸网络域名家族的检测设备，包括：
存储器，用于存储计算机程序；
处理器，用于执行所述计算机程序时实现上述任一种所述僵尸网络域名家族的检测方法的步骤。

15 又一方面，本申请还提供了一种计算机可读存储介质，所述计算机可读存储介质上存储有计算机程序，所述计算机程序被处理器执行时实现上述任一种所述僵尸网络域名家族的检测方法的步骤。

20 综上，本申请通过对域名之间各个维度的相似性进行分析，并将不同维度的相似性统一用图的数据结构进行表示，构成域名的时空关联图。在时空关联图上执行图计算算法，经过连通分支发现、社区发现、基于网页排名的异常点检测发现具有强内聚、弱外联性质的若干子图，每个子图对应一个僵尸网络域名家族。本申请并不只是仅仅依赖某一维度的特征对属于同一僵尸网络家族的 C&C 域名进行聚类与检测，而是综合考虑域名之间各个维度的相似性关系，因此能够对属于同一僵尸网络家族的域名更加全面的分析，并且具备更强的可扩展性，新增的维度的特征也可以很容易
25 加入到本申请提出的检测方法中去。本申请提出的检测方法具备很强的实时性，并不需要一定要获得 C&C 域名对应的恶意文件样本之后才能对其所属的僵尸网络家族进行分析。

本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其它实施例的不同之处，各个实施例之间相同或相似部分互相参见

即可。对于实施例公开的装置而言，由于其与实施例公开的方法相对应，所以描述的比较简单，相关之处参见方法部分说明即可。

5 专业人员还可以进一步意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

10 结合本文中所公开的实施例描述的方法或算法的步骤可以直接用硬件、处理器执行的软件模块，或者二者的结合来实施。软件模块可以置于随机存储器（RAM）、内存、只读存储器（ROM）、电可编程 ROM、电可擦除可编程 ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。

15 以上对本申请所提供的僵尸网络域名家族的检测方法、装置、设备以及计算机可读存储介质进行了详细介绍。本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想。应当指出，对于本技术领域的普通技术人员来说，在不脱离本申请原理的前提下，还可以对本申请进行若干改进和修饰，这
20 些改进和修饰也落入本申请权利要求的保护范围内。

权 利 要 求

- 1、一种僵尸网络域名家族的检测方法，其特征在于，包括：
获取可疑域名；
基于各所述可疑域名在不同维度之间的关联性，构建域名时空关联图；
5 在所述域名时空关联图中，各可疑域名作为一个节点，拥有至少一种关联性的两个域名之间构成一条边，两个域名之间的关联性作为边的属性值；
根据图计算中各个节点紧凑程度的判断指标，确定得到所述域名时空关联图中联系紧凑的域名，将对应域名的集合作为僵尸网络域名家族。
- 2、如权利要求 1 所述的僵尸网络域名家族的检测方法，其特征在于，
10 所述根据图计算中各个节点紧凑程度的判断指标，确定得到所述域名时空关联图中联系紧凑的域名包括：
根据稀疏度和平均聚类程度，确定得到所述域名时空关联图中联系紧凑的域名。
- 3、如权利要求 1 所述的僵尸网络域名家族的检测方法，其特征在于，
15 所述根据图计算中各个节点紧凑程度的判断指标，确定得到所述域名时空关联图中联系紧凑的域名包括：
对所述域名时空关联图中的异常子图以及异常节点进行去除，确定得到所述域名时空关联图中联系紧凑的域名。
- 4、如权利要求 3 所述的僵尸网络域名家族的检测方法，其特征在于，
20 所述对所述域名时空关联图中的异常子图以及异常节点进行去除包括：
依据连通性将所述域名时空关联图分解为多个连通分支构成的子图，去除不满足预设第一连通性指标的异常子图；
使用社区发现算法对去除异常子图后剩余的子图进行分割，得到多个连通分支构成的社区，去除不满足预设第二连通指标的异常社区；
25 采用网页排名算法度量每个节点在去除异常社区后剩余的社区中的重要程度，去除重要程度低的潜在异常点。
- 5、如权利要求 1 所述的僵尸网络域名家族的检测方法，其特征在于，所述不同维度之间的关联性包括以下特征中的任意组合：
域名文法特征上的相似性、域名与病毒关联上的相似性、域名解析出

的 IP 地址上的相似性、域名访问流量上的相似性。

6、如权利要求 1 至 5 任一项所述的僵尸网络域名家族的检测方法，其特征在于，在所述将对应域名的集合作为僵尸网络域名家族之后还包括：

5 对所述僵尸网络域名家族的家族特征进行分析，并根据分析得到的家族特征对应用场景进行安全检测。

7、如权利要求 6 所述的僵尸网络域名家族的检测方法，其特征在于，在所述将对应域名的集合作为僵尸网络域名家族之后还包括：

通过判断新域名与已知域名之间的关联性，确定新域名是否属于所述僵尸网络域名家族，以便对所述僵尸网络域名家族的变种和扩张进行监控。

10 8、如权利要求 1 所述的僵尸网络域名家族的检测方法，其特征在于，所述基于各所述可疑域名在不同维度之间的关联性，构建域名时空关联图，包括：

15 将各所述可疑域名在不同维度之间的关联性表示为三元组的形式，其中，所述三元组中的关联性指标包含关联的维度与在该维度上两个域名的相似性度量；

使用图数据库基于所述三元组构建域名时空关联图。

9、如权利要求 4 所述的僵尸网络域名家族的检测方法，其特征在于，所述第二连通指标比所述第一连通性指标在连通性指标上更加严格。

10、一种僵尸网络域名家族的检测装置，其特征在于，包括：

20 域名获取模块，用于获取可疑域名；

关联图构建模块，用于基于各所述可疑域名在不同维度之间的关联性，构建域名时空关联图；在所述域名时空关联图中，各可疑域名作为一个节点，拥有至少一种关联性的两个域名之间构成一条边，两个域名之间的关联性作为边的属性值；

25 检测模块，用于根据图计算中各个节点紧凑程度的判断指标，确定得到所述域名时空关联图中联系紧凑的域名，将对应域名的集合作为僵尸网络域名家族。

11、如权利要求 10 所述的僵尸网络域名家族的检测装置，其特征在于，所述检测模块用于：根据稀疏度和平均聚类程度，确定得到所述域名时空

关联图中联系紧凑的域名。

12、如权利要求 10 所述的僵尸网络域名家族的检测装置,其特征在于,所述检测模块用于: 对所述域名时空关联图中的异常子图以及异常节点进行去除, 确定得到所述域名时空关联图中联系紧凑的域名。

- 5 13、如权利要求 12 所述的僵尸网络域名家族的检测装置,其特征在于,所述检测模块用于: 依据连通性将所述域名时空关联图分解为多个连通分支构成的子图, 去除不满足预设第一连通性指标的异常子图; 使用社区发现算法对去除异常子图后剩余的子图进行分割, 得到多个连通分支构成的社区, 去除不满足预设第二连通指标的异常社区; 采用网页排名算法度量
10 每个节点在去除异常社区后剩余的社区中的重要程度, 去除重要程度低的潜在异常点。

- 14、如权利要求 10 所述的僵尸网络域名家族的检测装置,其特征在于,所述不同维度之间的关联性包括以下特征中的任意组合: 域名文法特征上的相似性、域名与病毒关联上的相似性、域名解析出的 IP 地址上的相似性、
15 域名访问流量上的相似性。

15、如权利要求 10 至 14 任一项所述的僵尸网络域名家族的检测装置,其特征在于,还包括:

- 安全检测模块,用于在将对应域名的集合作为僵尸网络域名家族之后,对所述僵尸网络域名家族的家族特征进行分析, 并根据分析得到的家族特征对应用场景进行安全检测。
20

16、如权利要求 15 所述的僵尸网络域名家族的检测装置,其特征在于,还包括:

- 监控模块,用于在将对应域名的集合作为僵尸网络域名家族之后,通过判断新域名与已知域名之间的关联性, 确定新域名是否属于所述僵尸网络域名家族, 以便对所述僵尸网络域名家族的变种和扩张进行监控。
25

17、如权利要求 10 所述的僵尸网络域名家族的检测装置,其特征在于,所述关联图构建模块, 具体用于将各所述可疑域名在不同维度之间的关联性表示为三元组的形式, 其中, 所述三元组中的关联性指标包含关联的维度与在该维度上两个域名的相似性度量; 使用图数据库基于所述三元组构

建域名时空关联图。

18、如权利要求 13 所述的僵尸网络域名家族的检测装置,其特征在于,所述第二连通指标比所述第一连通性指标在连通性指标上更加严格。

19、一种僵尸网络域名家族的检测设备,其特征在于,包括:

5 存储器,用于存储计算机程序;

处理器,用于执行所述计算机程序时实现如权利要求 1 至 9 任一项所述僵尸网络域名家族的检测方法的步骤。

20、一种计算机可读存储介质,其特征在于,所述计算机可读存储介质上存储有计算机程序,所述计算机程序被处理器执行时实现如权利要求
10 1 至 9 任一项所述僵尸网络域名家族的检测方法的步骤。